

Data Description Sheet

1. A description of which author(s) handled the data and conducted the analyses.

Ioannis Floros, Shane Johnson, and Wanjia Zhao were in charge of the data and conducted the analyses. Particularly, (1) Ioannis obtained companies' M&A information from Thomson One; (2) Wanjia obtained data from Compustat, Compustat Historical Segment File, CRSP, Kenneth French's data library, and directEDGAR; (3) Wanjia conducted analyses presented in Tables 1 – 10 as well as Tables OA1 – OA5 in the paper; (4) Shane Johnson and Ioannis Floros jointly hired a research assistant to hand collect the number of material contracts in firms' annual filings in our sample, and the research assistant was under the supervision of Ioannis Floros and Wanjia Zhao; (5) Ioannis Floros and Wanjia Zhao hand collected the number of redacted material contracts in firms' annual filings in our sample; (6) all authors were significantly involved in discussions related to the data collection, variable measurement, and empirical analyses.

2. A detailed description of how the raw data were obtained or generated, including data sources, the date(s) on which data were downloaded or obtained, and the instrument used to generate the data (e.g., for surveys or experiments). We recommend that more than one author is able to vouch for the stated source of the raw data.

Our data generation process included several rounds of revisions and updates. We describe below the final datasets that we used in the paper:

- 1) We obtained the Compustat historical segment data on March 18, 2021
- 2) We obtained all mandatory electronic filings covering our sample period from directEDGAR on December 28, 2021
- 3) We searched for key words including “confidential treatment” and “confidential request” in directEDGAR and obtained the mandatory electronic filings containing such key words on December 28, 2021
- 4) We searched and obtained from Thomson One M&A activities on January 21, 2019
- 5) We obtained Compustat annual data set to calculate covariates on December 28, 2021
- 6) We obtained CRSP data on July 2, 2021 to calculate portfolio returns and firm returns
- 7) We obtained data on three factors from Kenneth French's data library on April 30, 2021
- 8) We hired a research assistant to hand collect data on firms' number of material contracts and the number of redacted material contracts over the period from June 2023 to January 2024

3. If the data are obtained from an organization on a proprietary basis, the authors should privately provide the editors with contact information for a representative of the organization who can confirm data were obtained by the authors. The editors would not make this information publicly available. The authors should also provide information to the editors about the data sharing agreement with the organization (e.g., non-disclosure agreements, any restrictions imposed by the organization on the authors, such as restrictions to publish certain results).

The hand-collected data on firms' number of material contracts and the number of redacted material contracts is being used by the team in another research project, and therefore we consider

this data source to be proprietary at the moment. All other data collected are commercially available.

4. A complete description of the steps necessary to collect and process the data used in the final analyses reported in the paper. For experimental and survey papers, we require information about the instructions and instruments used to generate the data, subject eligibility and/or selection, as well as any exclusion criteria. The full set of instructions and instruments can be provided in the online appendix.

A complete description of the steps necessary to collect the data is detailed in #2. A complete description of the steps necessary to process the data is detailed in the program “20240904_data_clean_20260628_update.R”.

5. After downloading or obtaining the raw data, all manipulations of the data should be done via computer programs. The code for these manipulations should be included in the code submitted upon acceptance (see below). No manipulations of raw data can take place manually or outside the computer code provided. If compliance with this requirement is not feasible, the authors need to explain and disclose any manipulations of the raw data (e.g., manually created variables or file conversions). When feasible, we also encourage the authors to share the code that downloads the data.

Attached to this datasheet please find the replication package which includes (1) codes to be run to clean and manipulate data (“20240904_data_clean_20260628_update.R”) and (2) codes to be run to execute analyses (“20241013_tables_20260628_update.R”). They are both executed in R studio. All necessary packages to execute the program are self-contained and specified in the program.

6. The computer programs (i.e., code) used to (1) convert the raw data into the final dataset used in the analysis, (2) to execute the statistical or econometric analysis, and (3) to generate the tables or to produce the output used in constructing tables of the manuscript. A brief description that enables other researchers to understand and run the code should be provided. The purpose of this requirement is to facilitate replication and 3 to help other researchers understand in detail how the raw data were processed, the final sample was formed, variables were defined, outliers were treated, and which commands were used in the analysis, etc. This code or programming is in most circumstances not proprietary. However, we recognize that some parts of the code or data generation process may be proprietary, including from the authors’ perspective. Therefore, instead of disclosing the proprietary portion of the code or program, researchers can provide a detailed step-by-step description of the code or the relevant parts of the code such that it enables other researchers to arrive at the same results that the authors obtained and presented in their manuscript. In such cases, the authors should inform the editors upon initial submission, so that the editors can consider an exemption allowing the step-by-step description. Whenever feasible, authors are required to provide the identifiers (e.g., CIK, CUSIP) for their final sample. Authors should consult our FAQ Sheet on the JAR website for further details.

Please see the replication package. For sample construction, see

“20240904_data_clean_20260628_update.R”. For statistical analysis and output generation, see “20241013_tables_20260628_update.R”.

7. A comprehensive log file that shows the execution of the entire code. This log file should cover all the steps that convert the raw data into a final dataset and the execution of all statistical and econometric analyses presented in the tables of the manuscript. The portion of the log file that shows proprietary code or data may be masked. In this case, the reader should be referred to the step-by-step description provided as per the requirements in Item 6.

Please see the log files (“console_output_data_clean.txt” and “console_output_tables.txt”) that show the execution of the entire code. Note that the procedures of fetching Compustat and fetching CRSP data are embedded in the program “20240904_data_clean_20260628_update.R”. When these vendors maintain and update their databases, the outputs generated from our programs are slightly different from the version we generated and submitted in 2024, which is expected.

8. An assurance that the data and programs will be maintained by at least one author (usually the corresponding author) for at least six years, consistent with National Science Foundation guidelines.

Wanjia Zhao will maintain the data and programs according to the guideline.